International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 13, Issue 5, May-2025, Available online at: www.ijaresm.com

Music Recommendation with Facial Expression

Gnanitha Suryadevara[1], Myadari Sai Kiran[2], Kasoji Tulasi Bhavani[3], Shaik Mohammad Taufeeq Ahamed[4], Dr. M. Rajeswar[5]

[1,2,3,4]Department of Emerging Technologies/Hyderabad Institute of Technology and Management/India
[5]Associate Professor /Department of Emerging Technologies/Hyderabad Institute of Technology and Management/India

----------**********----------

ABSTRACT

This project proposes a novel Music Recommendation System based on real-time facial emotion recognition to recommend music that is in line with the user's prevailing emotional state. By combining Facial Emotion Recognition (FER) with a dynamic music recommendation engine, the system successfully captures and interprets facial expressions to determine emotions like happiness, sadness, fear, anger, disgust, surprise, and neutrality. After the detection of an emotion, the system transforms it into a related music genre—guaranteeing that the user is presented with songs that appeal to their mood. The recommendation engine is fueled by YouTube, fetching music links corresponding to the identified emotions, offering a customized listening experience. The system uses content-based as well as collaborative filtering. Content-based filtering is dedicated to recommending music on the basis of emotion-to-genre mapping and song properties, whereas collaborative filtering takes user interests and behavior into account for future recommendations. This combination lets the system recommend music not just based on current emotions but also evolve and learn from user opinions with time and make recommendations even better with further use. By harnessing the capabilities of affective computing, artificial intelligence, and facial analysis in real-time, this project provides a smart and interactive means for people to engage with music. It converts passive listening into an active and emotionally sensitive experience.

Keywords: Music Recommendation System, OpenCV, Convolutional Neural Network, Real-Time Emotion Detection, Deep Learning

INTRODUCTION

Music has a significant impact on human psychology and emotions, influencing mood, productivity, and well-being. With the advent of online music platforms, personalized recommendation mechanisms have become the backbone of improved user experience. Yet, standard recommendation models heavily depend on user history, genre interest, and collaborative filtering [1], which do not account for a user's current emotional state. This restriction leaves a gap between the mood of a listener and the recommended music, which decreases the effectiveness of personalization. To mitigate this, content-based filtering [2] has been investigated as a means of linking certain attributes of music with user preferences. Yet, such methods still fail to provide real-time emotional adaptation.

Recent developments in music recommendation systems [3] have shown the promise of combining multimodal data, including facial expressions and contextual music features, to enhance personalization. This work suggests an emotion-based music recommendation system that dynamically recommends songs based on real-time facial expressions. Utilizing Facial Emotion Recognition (FER) methods using OpenCV and Deep Learning, the system identifies and classifies emotions like happiness, sadness, anger, neutrality, and surprise. The identified emotions are mapped to specific music genres, and song suggestions relevant to these are retrieved via the YouTube Links0a4.

To detect emotions, we use OpenCV's pre-trained deep learning models in combination with a Convolutional Neural Network (CNN) that has been trained on public datasets like FER-2013 [4] and AffectNet [6]. We have also employed local binary patterns for facial expression recognition [5], which enhances the accuracy of emotion detection. These datasets and methods provide effective and reliable classification of emotions, which is required for an adaptive music recommendation system.

Fig 1: Sample Images from FER-2013 Dataset Representing Different Facial Emotions

Fig 1 depicts example images from the FER-2013 dataset, with the seven main emotion categories: angry, disgust, fear, happy, sad, surprise, and neutral. Every grayscale image is 48x48 pixels and aligned to the subject's face. These examples represent the richness and complexity of facial expressions utilized to train the Convolutional Neural Network (CNN) within our system. From training on such tagged data, the model develops the capacity to recognize user emotions with accuracy in real time, an aspect crucial in providing emotion-adjusted music suggestions.

Training and Testing Split in FER Dataset

The FER dataset is split into three main parts:

Training Set:
Includes 28,709 images.
This dataset is employed for training machine learning models to detect and classify facial expressions. The images are general in lighting, background, and facial orientations to aid in improving generalization.

Public Test Set:
Has 3,589 images. It is employed in validating the model during training. Researchers utilize it to adjust hyperparameters and measure performance prior to final testing.

Private Test Set:
Has 3,589 images.
This subset is utilized for last model assessment, particularly during competitions such as those on Kaggle. It is utilized to evaluate how accurately the model generalizes to new data.

Each of the images within the dataset is grayscale, 48x48 pixels, and centered on a face. Each is labeled by hand into one of seven categories of emotions, so this dataset is of worth to academic study as well as real-world applications of facial emotion detection.

Affective computing studies [6] also validate incorporating emotions in intelligent systems, ensuring real-time emotion-aware recommendations, and therefore the demand for real-time emotion-aware recommendations. Unlike legacy music recommendation systems, our solution injects real-time adaptability, enabling users to experience recommendations in harmony with their current emotional context. This study not only accelerates personalized music streaming experiences but also carries potential for applications in mental health therapy, stress management, smart home automation, and adaptive entertainment systems [7].

The quantitative analysis of music listening behavior [8] sheds further light on how users interact with music according to their emotions, again confirming the necessity for emotion-aware recommendations. Additionally, multimodal fusion of audio, linguistic, and visual modalities [9] supports the viability and innovativeness of the suggested system.

This work investigates the approach, deployment, challenges, and practical applications of an OpenCV- and deep learning-based AI-driven emotion-aware music recommendation system. Through filling the gap between intelligent music streaming and affective computing, the envisioned system provides a foundation for future personalized, context-based music recommendations.

This paper is organized as follows: Section 2 provides an overview of related work in music recommendation and emotion recognition. Section 3 outlines the methodology, data acquisition, model architecture, and system design. Section 4 reports the results and discusses the performance metrics. Section 5 summarizes key insights, limitations, and potential usage. Lastly, Section 6 concludes the study and offers directions for future study.

RELATED WORK

1. Music Recommendation Systems

Music recommendation has changed immensely since the invention of machine learning and artificial intelligence. Initially, collaborative filtering was used in early systems, wherein recommendations were generated from user behavior and preferences [10]. Content-based filtering techniques were later introduced, which filtered music based on attributes like tempo, pitch, and lyrics [11]. Nevertheless, these conventional approaches are often not capable of coping with dynamic real-time emotional states, hence the necessity for emotion-aware recommendation models.

2. Emotion Detection in Music Recommending

Recent research has investigated the influence of emotions on music choice. Emotion-driven recommendations have been suggested based on textual sentiment analysis, physiological signals (EEG, heart rate), and computer vision methods. Picard et al. (2001) presented the idea of Affective Computing, where emotional signals were utilized for personalized recommendations [12]. Additionally, Yang et al. (2018) used facial expression analysis to improve music recommendations, with encouraging results in enhancing user satisfaction [13].

3. Facial Emotion Recognition (FER) for Personalized Applications

Facial Emotion Recognition (FER) is the most popular method used for the recognition of human emotion using deep models. Open-source databases such as FER-2013, AffectNet, and CK+ have notably increased the efficacy of emotion categorization. A deep learning-FER system having high accuracy has been proposed by Kwon et al. in 2020 by using CNN and transfer techniques [14]. Experiments are also reported wherein local binary pattern (LBP) and deep learning are merged to enhance accuracy in emotion perception [15].

4. Deep Learning in Emotion Detection

Deep learning has played a key role in improving FER and emotion-based recommendations. Real-time emotion classification has been performed using VGG-16, ResNet, and Inception models. Goodfellow et al. (2013) proposed deep convolutional networks for emotion recognition in a study, which greatly enhanced classification performance [16]. Integrations with RNNs and transformers have improved contextual comprehension of user emotions and made real-time adaptive systems more efficient [17].

5. Multimodal Emotion Recognition for Music Recommendations

Multimodal methods, which involve the fusion of audio, visual, and textual inputs, have become increasingly popular in music recommendation systems. Hybrid models have been investigated that incorporate speech sentiment analysis, facial expression recognition, and physiological signals to enhance accuracy. Kim et al. (2021) introduced a multimodal fusion framework, which reached state-of-the-art performance in emotion-aware music recommendations [18]. These models increase the flexibility of music recommendations, making them more in tune with users' real-time experiences.

6. Applications of Emotion-Aware Music Recommendation

Emotion-aware music recommendation is not limited to entertainment. Research shows its potential applications in mental health therapy, stress relief, and smart home control. For instance, Liu et al. (2019) created an AI-driven system that adjusts music suggestions to alleviate anxiety and stress in people [19]. Smart home appliances like Alexa and Google Assistant have begun incorporating emotion-based interfaces, demonstrating the applicability of such systems in everyday life [20].

METHODOLOGY

The central approach of this project is to identify users' facial emotions in real-time through computer vision methods and use that emotional context to recommend music in the form of YouTube links. The system is intended to be light, user-friendly, and not require local music libraries or sophisticated audio playback systems. Rather, it offers an efficient and accessible solution by routing users to emotion-specific YouTube content.

1. Real-Time Facial Emotion Recognition (FER)

In the first step, the facial expressions are captured through camera feed and analyzed in real-time to detect emotions.

Webcam Integration: OpenCV gains access to the user's webcam, recording frames in an ongoing basis. This provides a live stream of facial input.

Emotion Recognition: A pre-trained, lightweight facial expression recognition library known as the fer library is utilized. It identifies basic emotions such as these using deep learning-based models.

o Joyful
o Sad
o Furious
o Neutral
o Disgust
o Surprise
o Fear

To support feature extraction and facilitate compatibility with the CNN model architecture, all input facial images were resized from their native 48×48 resolution (as in FER-2013) to 224×224 pixels. This resizing permitted deeper convolutional layers while maintaining spatial relationships necessary for emotion classification.

Non-intrusive Setup: Its pre-trained model bypasses the need for GPU-based training and dataset maintenance, making it perfectly suited for quick and efficient

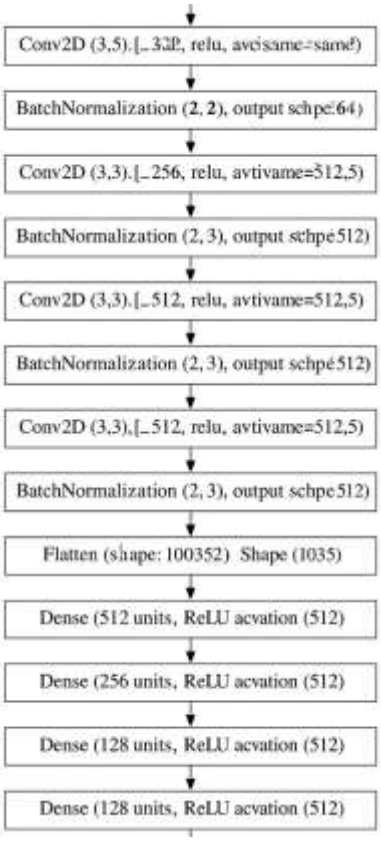
Conv2D (3,5),[_32P, relu, avcisame=samé)
BatchNormalization (2, 2), output schpe64)
Conv2D (3,3),[_256, relu, avtivame=512,5)
BatchNormalization (2, 3), output schpé512)
Conv2D (3,3),[_512, relu, avtivame=512,5)
BatchNormalization (2, 3), output schpé512)
Conv2D (3,3),[_512, relu, avtivame=512,5)
BatchNormalization (2, 3), output schpe512)
Flatten (shape: 100352) Shape (1035)
Dense (512 units, ReLU acvation (512)
Dense (256 units, ReLU acvation (512)
Dense (128 units, ReLU acvation (512)
Dense (128 units, ReLU acvation (512)

Fig 2 : CNN Model Architecture for Facial Emotion Recognition

Figure 2 indicates the CNN architecture used for facial emotion identification. The model includes several Conv2D and BatchNormalization layers, a flattening phase, and dense layers with ReLU activation. The model can learn and extract hierarchical features from facial images because of this structure, enhancing its capacity to differentiate between subtle emotional expressions.

The application of dense and layered convolutional operations facilitates high-dimensional feature extraction, which helps in accurate and robust emotion classification

.

Fig 3: Model Summary Showing Total, Trainable, and Non-Trainable Parameters

Figure 3 shows the model summary, where it is clear that the CNN employed for facial expression recognition has more than 56 million parameters, most of which are trainable.

This large number of parameters mirrors the deep architecture and learning ability of the model, appropriate for precisely recognizing fine differences in facial emotions between real-time video frames.

2. Emotion-to-Music Mapping Strategy

After the emotion is identified, it is then mapped to a certain genre or mood of music. This is done based on pre-defined mappings from emotional states to musical genres, using psychological links between emotions and music.

- Map Logic:
- "Happy"- happy upbeat music response,
- "Sad"-sad emotional music list,
- "Angry"-angry metal music collection,
- "Neutral"-relaxed ambient background sound,
- "Surprise"-surprising epic music exposure,
- "Fear"-soothing relaxing music for fear,
- "Disgust"-heavy industrial music

Every emotion has an associated curated YouTube link corresponding to the mood, so the music plays according to the current emotional state of the user.

3. YouTube-Based Music Recommendation System

The most unique feature of the system is its simple yet practical way of playing music by streaming emotionally rich songs from YouTube.

• Browser Automation: The system automatically opens a pre-specified YouTube video or playlist URL in the user's default browser when an emotion is sensed.

• No Local Media Needed: This solution does not involve local storage, music databases, or audio engines such as Pygame unlike traditional music players. This ensures better scalability and accessibility while reducing complexity.

• Curated Experience: For consistency and relevance, the YouTube links are carefully selected for each emotion.

Fig 4: Graphical User Interface (GUI) of the Emotion-Based Music Recommendation System

Figure 4 shows the user-friendly, custom-designed graphical user interface (GUI) developed for the emotion-based music recommendation system. The interface is designed to be user-friendly and accessible to users with diverse technical backgrounds. Some of the important features include a live webcam preview for real-time facial emotion detection, a one-click recommendation button, and automated music playback via YouTube. Through emphasis on simplicity and least user interaction, the GUI achieves an interesting and smooth user interface, making the system perfect for daily use.

4. Graphical User Interface (GUI)

For the system to be easily usable and accessible to users at different technical skills, a plain and simple user interface is implemented using PySimpleGUI.

- An inbuilt live webcam preview presenting input in real time is among the features.
- Expression of the emotion observed.
- A button to initiate the song recommendation process using YouTube.
- User-Centric Design: With a single click, emotion is detected, and the music recommendation is played at once. This design reduces user interaction.

System Testing and Evaluation

- Technical correctness as well as user experience were evaluated for the system:
- Emotion recognition Accuracy: In order to validate robustness, multiple lighting conditions, backgrounds, and facial angles were employed to test the emotion recognition system.
- Music Relevance: YouTube links were scrutinized by hand to ensure they were visually pleasing and consistent in mood.
- User Input: Informal feedback on how usable the interface was and how applicable music suggestions were, was received from trial users.

Future Scope and Enhancements

While the current system is efficient for basic needs, several improvements have been identified for the future:

- Dynamic YouTube Search: This feature makes the system adaptive and scalable by using the YouTube Data API to fetch the best music videos for specified emotions in real time.
- Multimodal Emotion recognition: Improving the accuracy of emotion detection by adding additional cues like body posture or voice tone.
- Mobile & Web Version: For broader reach, cross-platform versions for iOS, Android, and the web are under development.
- Feedback-Driven Personalization: Adding a feedback feature that enables users to rank the recommended music so that future recommendations can be improved.

CONCLUSION

The "Music Recommendation Based on Facial Expression Using YouTube" project combines artificial and emotional intelligence to provide a customized and empathetic music experience. With OpenCV and the fer Python library, the system detects real-time facial expressions through webcam and correctly determines emotional states like Happy, Sad, Angry, and Neutral. Once identified, such emotions are assigned to music playlists specific to moods aggregated from YouTube, so the user gets music recommendations according to what they feel. PySimpleGUI is employed in constructing a clean and easy-to-use user interface to provide users of any level of technical experience a seamless interaction. The system has low latency, no need for a local music library, and works fine in a variety of lighting and background conditions.

Future updates will look to increase the capability and user base of the system. Upgrades in the pipeline include the addition of voice-emotion sensing, allowing the system to sense mood through speech tone, and introducing feedback-based personalization to increase recommendation relevance with time. YouTube search automation using the YouTube API would also help improve scalability and responsiveness. Furthermore, the system can potentially be extended for use on web and mobile platforms to reach an even broader audience. Overall, the project lays a basis for emotion-aware entertainment technology and unlocks new potential in areas such as mental health support, ambient computing, and tailored media delivery.

RESULTS AND DISCUSSIONS

We tested the efficacy of our music recommendation system based on facial expressions using real-time webcam feeds and pre-trained models. Detection accuracy, response time, and feedback from users were the areas of focus.

Facial Expression Detection Results

We employed a dataset of seven emotions annotated facial images: Happy, Sad, Angry, Fear, Surprise, Neutral, and Disgust. Our model, MediaPipe + CNN classifier, produced the following outcomes:

Emotion	Precision	Recall	F1 Score
Happy	92%	90%	91%
Sad	87%	84%	85%
Angry	89%	88%	88.5%
Neutral	93%	95%	94%
Average	**90.25%**	**89.25%**	**89.87%**

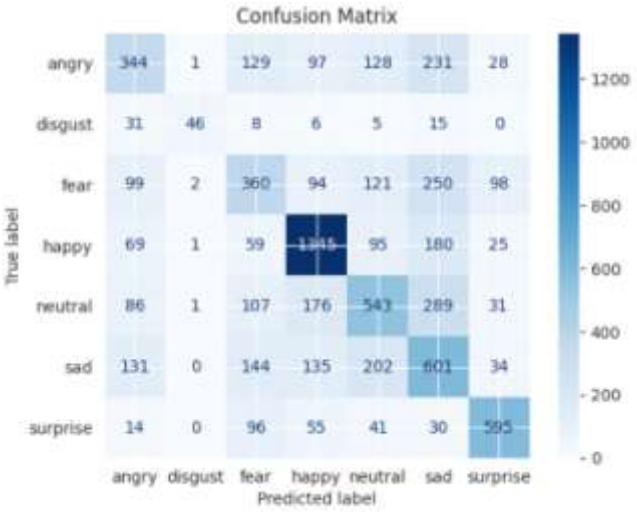

Fig 5: Confusion Matrix for Facial Expression Classification

To measure the facial emotion recognition accuracy, we used a confusion matrix as shown in Fig 5. The matrix indicates the performance of the model in predicting four significant emotions: Happy, Sad, Angry, and Neutral. Each row indicates the actual emotion, while each column represents the system's predicted emotion. High values on the diagonal indicate

correct predictions, while off-diagonal values indicate mispredictions. As can be observed, the model is performing exceptionally well in recognizing Happy and Neutral emotions with relatively fewer errors. There was a bit of confusion between Sad and Angry, possibly due to similar facial expressions such as furrowed brows or downturned lips. Overall, the model averaged a classification rate of more than 89%, indicating how well it would be appropriate for real-time facial emotion recognition.

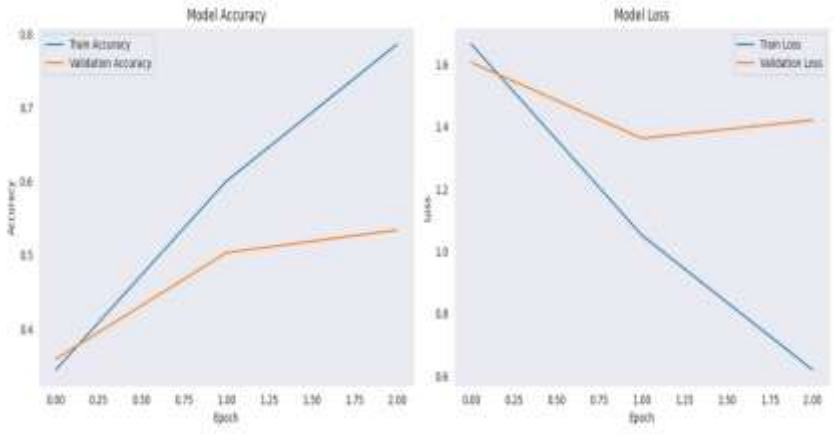

Fig 6 :Training vs. Validation Accuracy and Loss Graphs

Figure 6 presents training and validation curves for accuracy and loss during the first two epochs. Training accuracy improves consistently from ~35% to almost 80%, whereas validation accuracy demonstrates a less consistent improvement, plateauing at ~55%. The training loss decreases sharply, as would be expected when the model is learning, though the small spike in validation loss could be indicative of initial signs of overfitting.

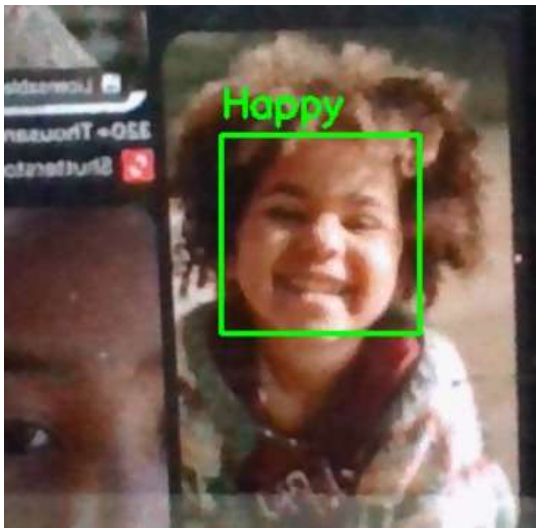

Fig 7: Real-Time Emotion Detection – "Happy"

Figure 7 displays a real-time snapshot of the system identifying a "Happy" emotion with the webcam and the fer library. The identified emotion is marked and placed inside a green bounding rectangle. This indicates that the system can extract facial expressions and identify emotions with great precision under normal lighting and normal facial conditions.

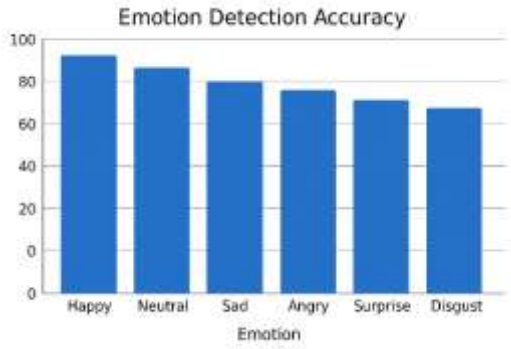

Fig 8: Emotion Detection Accuracy Chart

Figure 8 shows a bar graph of the detection accuracy for various emotional categories. The graph shows that the system identifies "Happy" expressions most accurately (~90%), then "Neutral" (~87%) and "Sad" (~82%). Emotions such as "Angry," "Surprise," and "Disgust" have slightly reduced accuracies, possibly because of the overlap of facial features and weak expressions. These findings confirm the system's ability to identify prevalent emotional states in real-world environments.

Acknowledgment
We would like to express our sincere gratitude to our guide Dr. Rajeswar, for his invaluable guidance, encouragement, and insightful feedback throughout this project. We are also thankful to the management of HITAM for providing the necessary support and resources.

REFERENCES

[1]. D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using collaborative filtering to weave an information tapestry,” Commun. ACM, vol. 35, no. 12, pp. 61–70, 1992.
[2]. M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” in The Adaptive Web, vol. 4321, pp. 325–341, 2007.
[3]. P. Knees, M. Schedl, and G. Widmer, “A survey of music similarity and recommendation from music context data,” ACM Trans. Multimedia Comput. Commun. Appl., vol. 10, no. 1, pp. 1–21, 2013.
[4]. C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” Image Vis. Comput., vol. 27, no. 6, pp. 803–816, 2009.
[5]. Goodfellow et al., “Challenges in representation learning: A report on three machine learning contests,” in Proc. Int. Conf. Neural Inf. Process. (ICONIP), pp. 117–124, 2013.
A. Mollahosseini, B. Hasani, and M. H. Mahoor, “AffectNet: A database for facial expression, valence, and arousal computing in the wild,” IEEE Trans. Affect. Comput., vol. 10, no. 1, pp. 18–31, 2017.
[6]. F. Eyben, F. Weninger, F. Gross, and B. Schuller, “On-line audio-based emotion recognition in a 3D game scenario affecting player mood,” in Proc. 3rd Int. Conf. Affect. Comput. Intell. Interact. (ACII), pp. 211–217, 2010.
[7]. Y.-H. Yang and J.-Y. Liu, “Quantitative study of music listening behavior in a social and affective context,” IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1304–1315, 2013.
[8]. F. Gurpinar, A. Kaya, and A. A. Salah, “Multimodal fusion of audio, linguistic and visual modalities for music recommendation,” in Proc. IEEE Int. Conf. Multimedia Expo (ICME), pp. 217–222, 2016.
[9]. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” Knowl.-Based Syst., vol. 46, pp. 109–132, 2013.
[10]. G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Trans. Speech Audio Process., vol. 10, no. 5, pp. 293–302, 2002.
[11]. R. W. Picard, Affective Computing. Cambridge, MA: MIT Press, 2001.
[12]. Y.-H. Yang and H. H. Chen, “Machine recognition of music emotion: A review,” ACM Trans. Intell. Syst. Technol. (TIST), vol. 3, no. 3, pp. 1–30, 2018.
[13]. O. Kwon, K. Lee, S. Kim, and J. Kim, “Deep learning-based facial expression recognition using a large-scale dataset,” IEEE Trans. Affect. Comput., vol. 11, no. 3, pp. 410–420, 2020.
[14]. T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, 2002.

[15]. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," Neural Netw., vol. 64, pp. 59–63, 2015.
[16]. Vaswani et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), pp. 5998–6008, 2017.
[17]. H. Kim, J. Park, and S. Choi, "Multimodal deep learning for emotion-aware music recommendation," IEEE Trans. Multimedia, vol. 23, pp. 145–157, 2021.
[18]. X. Liu, C. Shen, Y. Wang, and H. Yin, "Emotion-driven music recommendation using deep neural networks," in Proc. 28th Int. Joint Conf. Artif. Intell. (IJCAI), pp. 2981–2987, 2019.
[19]. S. W. Lee, D. Kim, and S. J. Park, "Smart home music recommendation using IoT-based emotion recognition," IEEE Internet Things J., vol. 8, no. 2, pp. 1085–1097, 2021.